DOCUMENT RESUME

ED 410 878 HE 030 475

AUTHOR Stanley, Elizabeth C.; Patrick, William J.
TITLE Teaching and Research Quality Indicators and the Shaping of Higher Education. AIR 1997 Annual Forum Paper.
PUB DATE 1997-05-00
NOTE 25p.; Paper presented at the Annual Forum of the Association for Institutional Research (37th, Orlando, FL, May 18-21, 1997).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Accreditation (Institutions); Cluster Analysis; College Outcomes Assessment; Comparative Education; Cross Cultural Studies; *Educational Quality; Financial Support; Foreign Countries; Government Role; Government School Relationship; *Higher Education; Institutional Evaluation; Qualitative Research; Research; Teacher Effectiveness
IDENTIFIERS *AIR Forum; Performance Indicators; Ranking; *United Kingdom; United States

ABSTRACT

Two important sets of performance indicators for institutions of higher education have become established in the United Kingdom: research quality ratings and teaching quality ratings. The research quality ratings and, to a lesser extent, the teaching quality ratings influence the level of government funding provided to higher education institutions. This paper considers the correlations between the two ratings and the possible consequences of policies which reshape the higher education sector by concentrating research resources in a limited number of institutions. Comparisons are made between quality assurance/assessment approaches in the United Kingdom and the United States, finding that U.S. higher education is much larger, more heterogeneous and has less government control than U.K. higher education, While the U.S. system of colleges and universities is generally unranked (by those responsible for accreditation), the UK system includes rankings. Use of use various analytical approaches to compare teaching and research ratings for both systems concluded that it remains unclear whether the measurement standards will lead to improvements in teaching and research. (Contains 38 references.) (Author/DM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 410 878

Teaching and Research Quality Indicators and the Shaping of Higher Education

William J. Patrick
Planning Officer
University of Glasgow
University Avenue
Glasgow G12 8QQ
Scotland
44-141-330-4218
Fax: 44-141-330-4920
E-mail: bill@mis.gla.ac.uk

Elizabeth C. Stanley
Director of Institutional Research
Iowa State University
215 Beardshear Hall
Ames, IA 50011-2030
515-294-1181
Fax: 515-294-4835
E-mail: estanley@iastate.edu

Paper to be presented at the 37th Annual Forum of
The Association for Institutional Research
Orlando, Florida
May 18-21, 1997

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

AIR

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

IAE 030425

for Management Research, Policy Analysis, and Planning

This paper was presented at the Thirty-Seventh Annual Forum of the Association for Institutional Research held in Orlando, Florida, May 18-21, 1997. This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of Forum Papers.

Jean Endo
Editor
AIR Forum Publications

Abstract

Two important sets of performance indicators have become established in the United Kingdom - research quality ratings and teaching quality ratings. The research quality ratings and, to a lesser extent, the teaching quality ratings, influence the level of government funding provided to higher education institutions . This paper will consider the correlation between the two ratings and the possible consequences of policies which reshape the higher education sector by concentrating research resources in a limited number of institutions. Comparisons will be made between quality assurance/assessment approaches in the United Kingdom and those in the United States.

Introduction

Increasing accountability for the use of public funding is well recognized as one of the prime influences on institutional development in higher education on both sides of the Atlantic. Both accountability demands and interests in institutional improvement have led to an increased focus on various aspects of quality. Quality assessments in the United Kingdom have been used to increase selectivity in the distribution of resources in support of research and have created a heightened awareness of quality issues in teaching and learning. The higher education systems in the United Kingdom (U.K.) and the United States (U.S.) differ markedly, insofar as the latter is much larger, more heterogeneous, and subject to fewer central governmental controls. One of the consequences of greater regulation on the part of the U.K. authorities is that a substantial amount of statistical information has been produced, and it is possible to speculate on the lessons to be learned from an examination of this material for the future shaping of higher education in the U.K. and the U.S.

Measures of Quality in Higher Education

In the United Kingdom, the Further and Higher Education Acts of 1992 brought together under unitary regional higher education funding councils (one for each of England, Scotland and Wales) the so-called "old" universities, the "new" universities, and the higher education colleges, as well as a small number of other institutions. (Up until that time, research had been conducted mainly in the "old" universities. The "new" universities had previously been polytechnics. The higher education colleges, which do not have degree-awarding status, fall into two broad groups: those which offer a range of courses, usually narrower than the universities, and those which specialize in just one subject. The "other" institutions also include the academies of music and drama.) The Acts also brought into clear focus the importance of quality in teaching and learning in U.K. higher education. A new vocabulary was created. In the teaching and learning area, this includes Quality Assurance, Quality Control, Quality Audit, and Quality Assessment.

Quality Assurance encompasses all the policies, systems, and processes directed towards ensuring the maintenance and enhancement of the quality of teaching and learning in higher education. Under this general umbrella, Quality Control relates to the arrangements (procedures, standards, and organization) within higher education institutions which verify that teaching and assessment are carried out in a satisfactory manner. These would normally include the external examiner system.

Quality Audit is the process of ensuring that the quality control arrangements in an institution are satisfactory. In practice, the prime responsibility for quality audit lies individually or collectively with institutions. It extends to the totality of quality assurance in an institution and may include staff development and curriculum design. External quality control has been conducted in the U.K. by the Quality Audit Unit of the Higher Education Quality Council which is collectively owned and funded by the institutions, and which was established in 1992.

Quality Assessment is the process of external evaluation of the quality of teaching and learning in higher education. Since 1992, responsibility for quality assessment has been a statutory obligation of the higher education funding councils. The external assessment by peers of the actual provision of education in particular subjects is carried out by scrutiny of institutional documentation and student work, direct observation, interview, and by reference to performance indicators such as completion rates. Although the external assessment mechanisms have been used broadly to scrutinize the inputs, processes, and outputs associated with educational processes, it is felt that outputs may receive less attention than inputs. A significant number of assessment ratings have been published and will be considered in this paper.

After five years of rapid development, the U.K. systems of external scrutiny of teaching and learning are poised to change significantly again. A single Quality Assurance Agency was introduced early in 1997, in order to combine and simplify the external Quality Assessment and Quality Audit procedures, and to address mounting concern over the educational standards attained by graduates in an increasingly diverse system. The traditional external examiner system is no longer felt to be effective in maintaining consistent standards in the face of the huge expansion in recent years in the numbers of courses and undergraduates. For the first time, a single agency will be responsible for both quality and standards (Joint Planning Group for Quality Assurance in Higher Education [JPG], 1996a, 1996b).

The United States has several parallel assessment processes, although the distinctions between quality assurance, control, audit, and assessment are not as specific. Institutional accreditation, provided primarily by regional associations of colleges and universities, and specialized accreditation of specific program areas, carried out by professional associations, provide evidence that institutions or programs meet specified minimum requirements and criteria, and encourage institutional quality improvement. Accreditation processes normally require an institutional self-study, and emphasize peer review, but do not lead to a ranking of institutions or programs. The rankings published by various magazines and college guides are typically based on the evaluation of institutionally reported data and the application of ranking criteria designed by the publishers.

In the United States, recent interest has focused on the assessment of student academic achievement as an important outcome measure and a tool for program improvement. Accrediting agencies and a number of states require that institutions implement programs for the assessment of student outcomes, and it is expected that the results of these assessments will be used to enhance quality. Accreditation processes have been criticized in the past for relying too strongly on inputs; the emphasis on outcomes assessment thus represents a significant change.

State and federal agencies may require institutions to report other measures of institutional effectiveness, but do not generally conduct quality assessments of the type developed in the U.K. At the program level, the practice of program review, often involving a departmental or unit self-study and external review, provides an institutional process for quality review.

Research quality in higher education institutions in the U.K. has been evaluated through four national research assessment exercises (RAEs): in 1986, 1989, 1992 and 1996. These exercises have assumed very great importance, partly because of the reputational significance of the outcomes, and partly because these outcomes have

determined to a considerable and increasing extent the levels of central governmental funding for research. In each exercise, institutions have submitted standard statistical and narrative material to central, specialist panels for peer-review assessment. The information considered by the panels has been purely documentary. Site visits have not formed part of the exercises. Reputational ratings have not been used.

In the United States, studies of research-doctorate programs published in 1966, 1970, 1982 and 1995 rated doctorate programs on the basis of reputational measures and, in the latter two studies, a number of other quantitative factors. The 1995 report provides rankings of programs by the "scholarly quality of program faculty" (Goldberger, Maher, and Flattau, 1995). There is not, however, the direct linkage to government funding for research which exists in the U.K.

Quality Measures in Teaching and Learning

This section focuses on the assessment of teaching quality in England and Scotland. (The Welsh funding council established a collaborative assessment program with the English funding council, although there have been a number of variations in the approaches adopted for assessment in Wales. The English funding council manages quality assurance arrangements for the two universities in Northern Ireland.)

Quality assessment in England has moved through two distinct phases. Between the autumn of 1992 and the summer of 1995, the provision of education in 15 subjects was assessed, using a three-point scale of "excellent," "satisfactory," and "unsatisfactory" quality of education. The factors taken into account included: the breadth of teaching; learning and assessment activities; students' achievements; the curriculum; the application of learning resources (such as library, equipment, information technology and laboratories); student support and guidance; and academic management at the subject level. An important feature was that quality was measured not against any absolute standards, but rather against the aims and objectives set by institutions themselves. The model was one of "fitness for purpose" with the objectives being specified by the providers. Consequently, it is not appropriate to compare results between institutions having widely differing missions. The process was one of self-assessment, followed by peer review which, in some cases, included a site visit. No judgment of "excellent" or "unsatisfactory" was made without a site visit. In all, 972 assessments were made. The Higher Education Funding Council for England reported, "With one exception, the proportion of excellent education by subject ranges from 10 percent to 49 percent of providers. Those subjects where peers found excellent quality provision in fewer than 20 percent of providers were in science, engineering and technology disciplines...." The Council also observed some differentiation in the proportion of excellent ratings between subjects and between the old universities and the other types of institution (Higher Education Funding Council for England [HEFCE], 1995). Figure 1 provides comparisons for a total of 408 assessments analyzed for the old universities, 356 for the new universities, and 200 for other institutions. The "other institutions" group includes the higher education colleges and some specialized institutions.

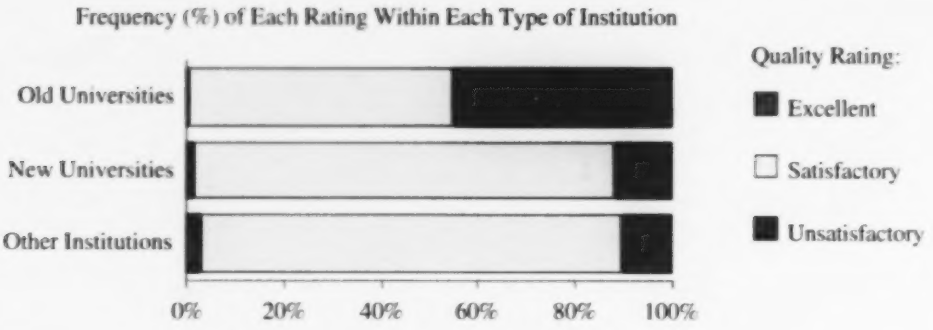

Fig. 1. Teaching Quality Assessments in England 1992-95

The English system of quality assessment was changed in April 1995, after widespread consultation. Site visits became universal rather than selective. Six aspects of provision were each graded on a four-point assessment scale (1 - 4), making a total attainable score of 24. These six different aspects of provision were: curriculum design, content and organization; teaching, learning and assessment; student progression and achievement; student support and guidance; learning resources; and quality assurance and enhancement (HEFCE, 1994). The scale points were defined as:

4 This aspect makes a full contribution to the attainment of the stated objectives; the aims set by the subject provider are met.

3 This aspect makes a substantial contribution to the attainment of the stated objectives; however, there is scope for improvement; the aims set by the subject provider are met.

2 This aspect makes an acceptable contribution to the attainment of the stated objectives, but significant improvement could be made; the aims set by the subject provider are broadly met.

1 The aim and/or objectives set by the provider are not met; there are major shortcomings that must be rectified.

Eight subjects were assessed in 1995-96, and a total of 272 visits were made. Ninety-nine percent of provision was quality-approved. Forty-two percent of all ratings were fours, and fifty percent were threes. Figure 2 shows the frequency of ratings on each of the quality aspects according to type of institution.

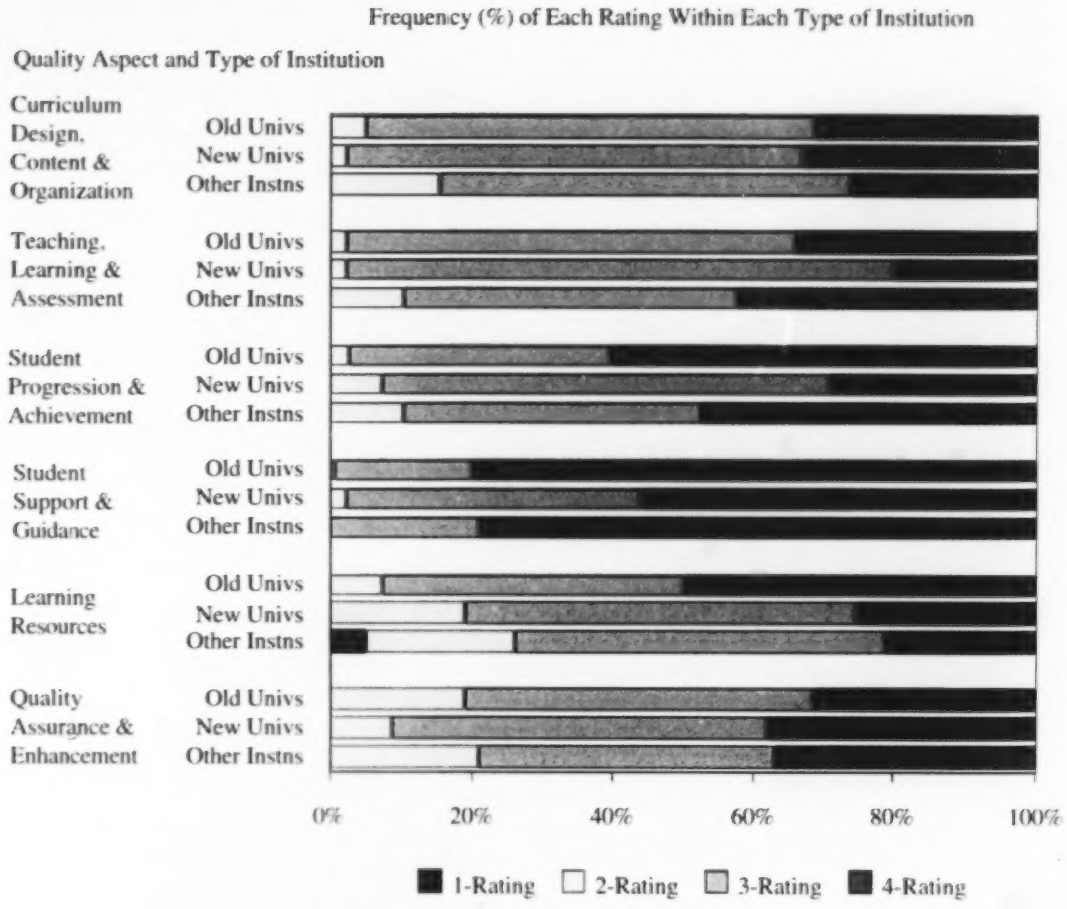

Fig. 2. Teaching Quality Assessments in England 1995-96

The results cannot be compared directly with the results obtained under the former single scale of "excellent," "satisfactory," and "unsatisfactory." There was no summative equivalent of an "excellent" rating under the new profiling method. Aggregating the six profile grades does not provide meaningful comparative information (HEFCE, 1997). However, the profiles identify the potential for improvement for all aspects graded 3 or below and, as such, constitute a more constructive tool for promoting improvement than did the previous summative ratings. Grade 2 scores were concentrated in the learning resources aspect and in quality assurance and enhancement.

The profiles provide evidence of some resource-related difficulties in some of the new universities and the other institutions—for example, in staffing, equipment, information technology and library provision. They also point to weakness in some institutions in subject-level quality assurance and enhancement—for example, over-reliance on informal systems, and limited responsiveness once problems have been identified.

With the exception of the learning resources aspect and, to some extent, the student progression and achievement aspect, the profiling method distinguishes less strongly between the old universities and the other institutions than the summative grading approach, which had been used from 1992 to 1995, and the English funding council questioned the continuing validity of general comparisons based on the earlier approach (HEFCE, 1997). The Funding Council also observed some clear differences in the profiles of different subject areas. To date, there have been no differentials in the levels of funding in England to reflect differing teaching quality assessment results, although the possibility has received serious consideration.

Teaching and learning at each of Scotland's 21 institutions of higher education is being assessed over a five-year cycle by the Scottish Higher Education Funding Council (SHEFC), the central body responsible for the sector. The process is based on a self-assessment, followed by a site visit, typically lasting several days. The visiting teams are made up primarily of practitioners within the relevant disciplines, along with representatives of business interests and of the Council itself. Thirty-one subject assessments had been completed by the end of 1996. As in the English system, assessments measure the achievements of teaching and learning against institutional objectives, rather than against any absolute standards. Generally, eleven aspects of each area are assessed: aims and curricula; curriculum design and review; the teaching and learning environment; staff resources; learning resources; course organization; teaching and learning practice; student support; assessment and monitoring; students' work; output, outcomes and quality control. Each of these aspects is classified as "excellent," "highly satisfactory," "satisfactory," or "unsatisfactory." In order for an institution's teaching in a particular subject to be rated "excellent," at least seven of its aspects must be excellent. Figure 3 illustrates the frequency of ratings by type of institution.

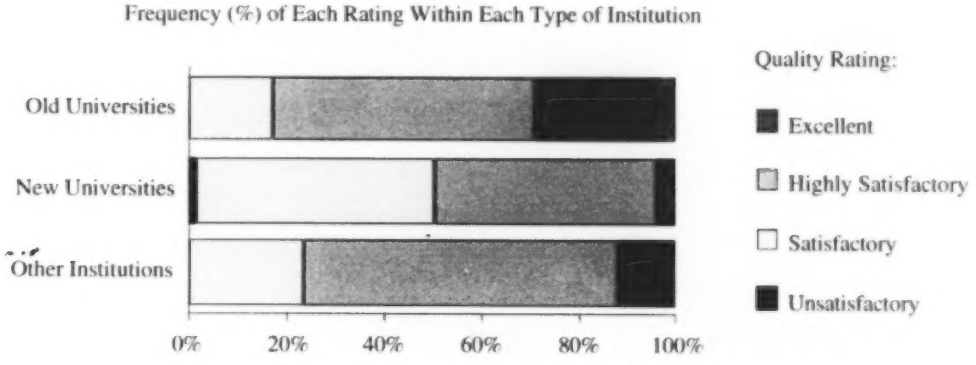

Fig. 3. Teaching Quality Assessments in Scotland 1992-96

The SHEFC has not sought to impose any uniform standards on the assessment teams working in different subject areas, and ratings have tended to vary considerably between them. Summary reports are distributed widely by SHEFC to high schools, and it is expected that this will influence student choice and increase competition (Scottish Higher Education Funding Council [SHEFC], 1996). The teaching ratings do have a financial impact in Scotland,

since departments which receive the accolade of "excellent" receive a five percent increase in governmental funding. Departments rated "unsatisfactory" are revisited in 12 months. If they are again rated "unsatisfactory," direct funding is withdrawn.

The U.S. does not have a similar rating system although institutional and specialized accreditation processes, as well as internal program review, examine the quality of educational programs in determining whether a program or institution meets minimum standards and is accomplishing its purposes. Like the U.K. quality assessment processes, accreditation involves an institutional self-study, followed by peer review and a site visit. The requirements and criteria of the six regional accrediting associations vary, but the North Central Association, for example, requires that institutions meet 24 General Institutional Requirements. These are concerned with mission, authorization, governance, faculty, educational programs, finances, and public information. Institutions must also satisfy five Criteria for Accreditation, including a criterion stating that "The institution is accomplishing its educational and other purposes." (North Central Association, 1994). Institutional accreditation evaluates the institution in terms of its mission, and recognizes that the diversity among institutional mission and purposes must be considered in the interpretation of the criteria. Institutional accreditation is necessary for eligibility for some forms of federal funding.

Existing rankings of undergraduate programs or institutions in the U.S. are primarily those produced by magazine or college guide publishers, such as U.S. News and World Report, Money magazine, and others. These are based primarily on the analysis of institutionally-supplied data, although reputational ratings may be included. There is considerable lack of standardization in the criteria employed, both between publishers and from year to year for a single publisher. These ranking are not considered, by institutions, to represent effective evaluations of the educational programs, but they may influence public opinion and student choice.

Quality Measures in Research

1996 U.K. Research Assessment Exercise

In the 1996 U.K. Research Assessment Exercise (RAE), each institution could make a submission in any of 69 units of assessment, defined to cover a full range of academic disciplines, including clinical medicine and dentistry (Higher Education Funding Councils, 1994b). Since the inception of the assessment exercises, panels have come to place increased emphasis on their appraisal of the research publications presented by faculty as their best work during the relevant assessment period (Higher Education Funding Councils, 1995). The adjudged quality of these publications, along with the information provided by institutions, has formed the basis of the quality ratings which in 1996 were awarded on a seven-point, criterion-referenced scale, developed from the earlier, five-point scale which had been used in 1989 and 1992 (Higher Education Funding Councils, 1994a):

> 5* (new in 1996; previously a subset of 5)—Research quality that equates to attainable levels of international excellence in a majority of sub-areas of activity and attainable levels of national excellence in all others.

5 (as before)—Research quality that equates to attainable levels of international excellence in some sub-areas of activity and to attainable levels of national excellence in virtually all others.

4 (as before)—Research quality that equates to attainable levels of national excellence in virtually all sub-areas of activity, possibly showing some evidence of international excellence, or to international levels in some and at least national levels in a majority.

3a (new in 1996; previously combined with 3b)—Research quality that equates to attainable levels of national excellence in a substantial majority of the sub-areas of activity, or to international level in some and to national level in others together comprising a majority.

3b (new in 1996; previously combined with 3a)—Research quality that equates to attainable levels of national excellence in the majority of sub-areas of activity.

2 (as before)—Research quality that equates to attainable levels of national excellence in up to half of the sub-areas of activity.

1 (as before)—Research quality that equates to attainable levels of national excellence in none, or virtually none, of the sub-areas of activity.

The average overall quality rating improved from 3.42 in 1992 to 3.73 in 1996 (assuming that 3a can be combined with 3b, and 5* with 5 to create a comparable 5-point scale) (Universities Funding Council, 1992; Higher Education Funding Councils, 1996). The total number of assessments increased by 6% from 2,730 to 2,893. The total number of faculty whose research was assessed rose by 11.2% from 43,211 to 48,068 (measured in full-time equivalents), while the overall number of faculty, including those whose research was not assessed, rose by about 18% from approximately 58,200 to 68,700. The percentage of research-active faculty declined from 74% to 70%.

Figure 4 illustrates the great variation in average quality ratings among units of assessment, both in 1992 and 1996. Biochemistry and Mineral & Mining Engineering have the highest average scores in 1996 (4.69 and 4.47 respectively), while Other Studies & Professions Allied to Medicine and Nursing continued to have the lowest average scores (2.97 and 2.39 respectively).

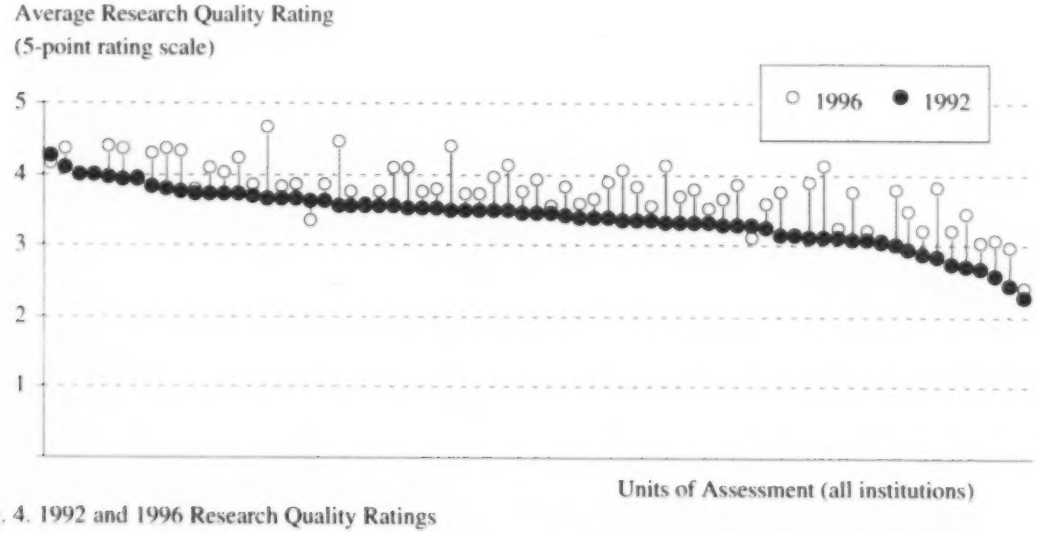

Fig. 4. 1992 and 1996 Research Quality Ratings

Those units of assessment exhibiting a reduction in the number of research-active faculty tended to have higher research ratings in 1996 than in 1992. The number of research-active faculty in the Arts and Humanities subjects increased, whereas in Science and Engineering subjects, such as Biochemistry, Chemistry, Mathematics, and Electrical and Electronic Engineering, decreases were experienced. The biggest increases in research-active faculty numbers were in Art and Design (+602 FTEs), Education (+492 FTEs), and Business and Management Studies (+286 FTEs). It is not clear whether these changes represent tactical maneuvering on the part of institutions, or whether they represent underlying changes in the pattern of research. Growth was most prevalent in the new universities.

Figure 5 provides examples of the relationships which exist between quality and quantity within units of assessment. In Chemistry, the relationship between quality and size is linear to a remarkable extent; in English relatively small departments have a wide range of results, while large departments achieve consistently high quality ratings.

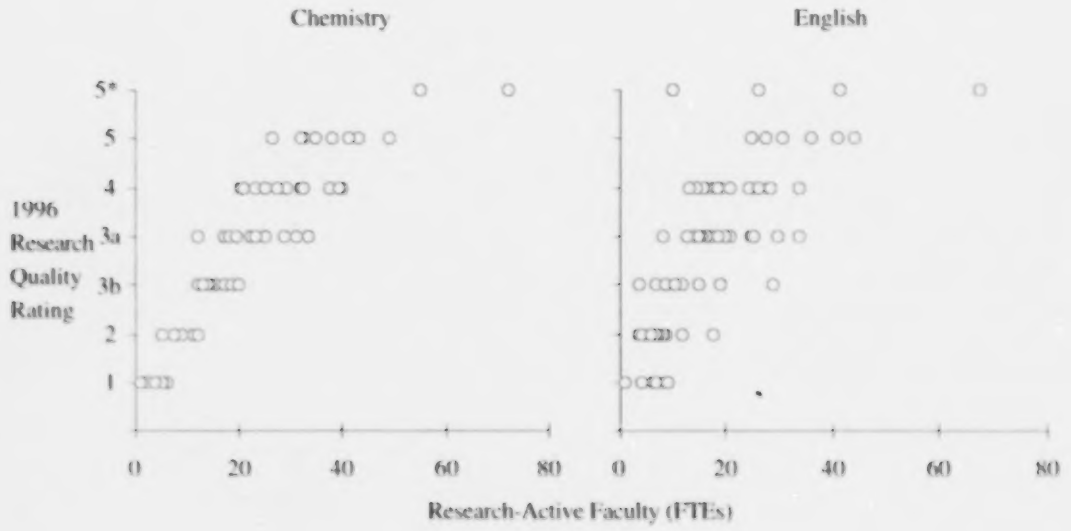

Fig. 5. 1996 Research Assessment Exercise—Quality and Size

This phenomenon exists not just in Chemistry and English but across a range of disciplines, as illustrated in Table 1.

Table 1. Correlations between Quality Ratings and Research-Active Faculty FTEs

Unit of Assessment	Spearman Rank Correlation Coefficients 1992 RAE	1996 RAE
Chemistry	0.82	0.91
English Language and Literature	0.82	0.80
Physics	0.76	0.79
Geography	0.71	0.78
Electronics & Electrical Engineering	0.67	0.74
Archaeology	0.66	0.74
Law	0.77	0.73
Theology	0.78	0.69

There was considerable divergence in the quality ratings for the different types of institution. Figure 6 shows that the old universities have a greater preponderance of high-quality units of assessment than do the other types of institution. Similar differences were observed in the 1992 Research Assessment Exercise results (Patrick and Stanley, 1996).

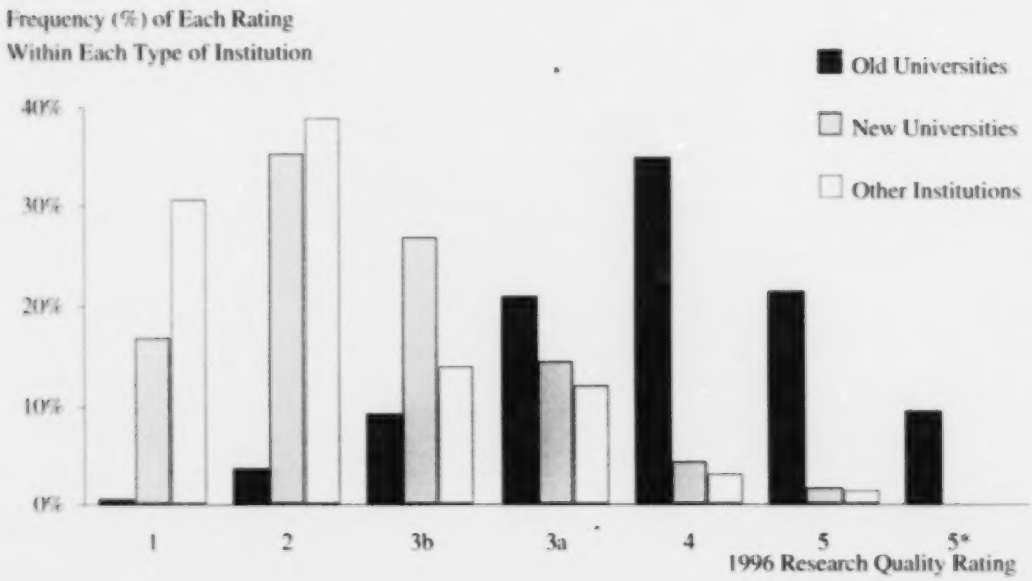

Fig. 6. 1996 Research Quality Ratings

The quality ratings have a significant influence on governmental research funding, and differences in funding levels according to quality ratings have increased in recent years. Figure 7 shows that in 1991/92, for example, the funding council's relative funding for departments in England increased linearly from 0.2 to 1 in the same proportion as the quality ratings themselves increased; by 1997/98, selectivity in the funds allocated by the funding council had increased to the extent that a unit of assessment having a quality rating of 1 or 2 would receive no funding at all; other quality ratings would attract funding on a scale of 0.30 for 3b, 0.44 for 3a, 0.67 for 4, 1 for 5, and 1.2 for 5*.

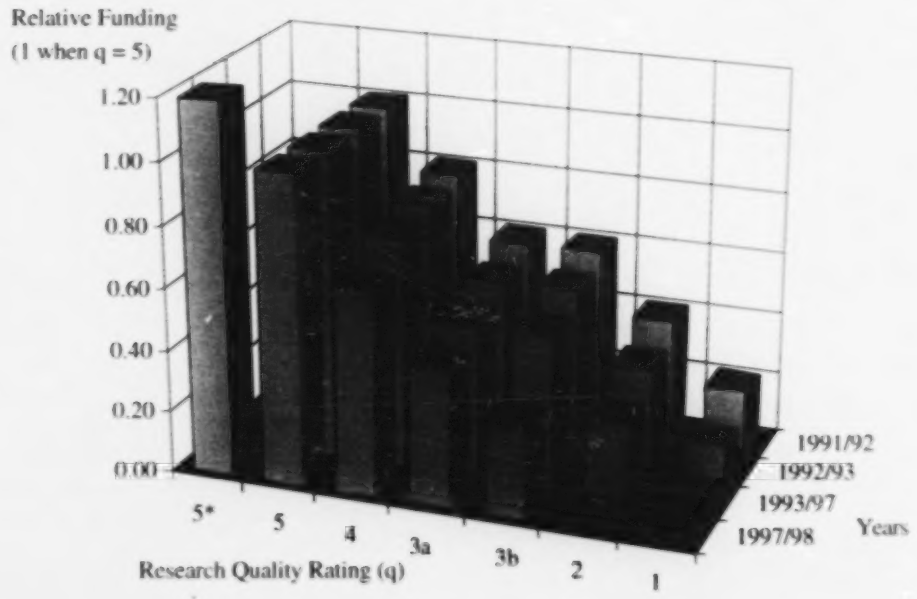

Fig. 7. Selectivity in Research Funding from 1991/92 to 1997/98 in England

Statistics collected nationally each year until 1993/94 by the Universities Statistical Record (USR) were used to search for relationships between research quality ratings and levels of departmental expenditure in the old universities. (See Universities Statistical Record, 1995, for example.) The USR published financial, personnel and student data for each of 37 academic cost centers, and it is possible to assign each of the 72 units of assessment used for the 1992 Research Assessment Exercise to one of these cost centers. There were very few significant correlations within the 37 academic cost centers between research quality ratings and levels of departmental spending of general funds per faculty member. No consistent correlations were found between research quality ratings and the faculty-student ratios within the cost centers (each representing a homogeneous group of academic subjects). Court (1996) has previously reported negative correlations between RAE scores and faculty-student ratios but his analysis was at the institutional level and may reflect their varying subject mixes, as well as differing levels of institutional infrastructure and support.

Other variables which were found to have influenced significantly the 1992 Research Assessment Exercise quality ratings were the numbers of articles in academic journals, total external research income, postgraduate research students, short works, and books (all divided by the number of researchers). The influence of the variables differed by subject area (Patrick and Stanley, 1996).

1993 U.S. Study of Research-Doctorate Programs

The 1993 study of research-doctorate programs included 3,634 programs in 41 fields at 274 U.S. universities (Goldberger, Maher, and Flattau, 1995). The study was conducted by a committee appointed by the National Research Council (NRC) at the request of the Conference Board of Associated Research Councils. It was designed to build upon and update a similar 1982 study (Jones, Lindzey, and Coggeshall, 1982). The committee conducted a National Survey of Graduate Faculty to collect reputational measures on two dimensions of program quality—the "scholarly quality of program faculty" and the program's "effectiveness in educating research scholars/scientists." The mean ratings of the "scholarly quality of program faculty" were used to rank order programs within each of the 41 fields in the study. Quality groupings were then subdivided into quartiles to facilitate discussions of program characteristics and to discourage overemphasis on small and possibly insignificant differences between closely ranked programs. The quality ratings were also expressed as standard scores to compensate for differences in the mean ratings between disciplines.

Data on the characteristics of program faculty, students, and doctoral graduates were obtained from several national data bases. These included publication and citation measures, data on federal research support, faculty honors and awards, enrollments and graduates, and the characteristics of doctoral recipients. Although the accuracy and comprehensiveness of some of the data were dependent on the accuracy of faculty lists provided by the institutions, the analysis of these data permit the correlation of program quality with other program characteristics. It was shown, for example, that the top-rated programs in most fields tend to have more faculty and more graduate students than the lower-rated programs. This is comparable to the finding reported here for the U.K. research assessments. Most of the research-doctorate programs received some type of federal support for research. A review of publications and citations for faculty in Sciences and Engineering indicated that faculty in highly-rated programs are cited more frequently than those in lower-rated programs. In the Arts and Humanities, faculty in the top-rated programs are more likely to have received honors and awards. The extensive data base developed for the study has been made available to other researchers for continued analysis, and further studies of the data base are expected to demonstrate additional relationships.

It has been noted for both the U.K. research assessments (Taylor, 1995) and the U.S. rankings of research-doctorate programs (Maher, 1996) that it is important to analyze the results by program, rather than for total institutions. Webster and Skinner, who presented global institutional comparisons by calculating the mean scores of all ranked programs within institutions (Webster and Skinner, 1996a) responded that the aggregation of rankings of sub-areas into global rankings is a frequently-used approach, and is presumably useful (Webster and Skinner, 1996b). It has also been observed, for both departments and institutions, that the group of top-ranked units tends to be quite

BEST COPY AVAILABLE

stable over time (Bensman, Hamaker and Wilder, 1997; Fogel, 1996). Although it is possible for new units to move into the most highly-ranked groups (Graham and Diamond, 1997), the probability of such movement is small. This is, in part, attributed to the "double-edged Matthew Effect" in which success is rewarded by increased chance of success, and failure is punished by the increased chance of further failure (Bensman et al., 1997).

Correlations Between Research Quality and Teaching Quality

Assuming that the research quality and teaching quality ratings are valid indicators of quality in their respective areas, it is then of interest to examine the relationship between them, particularly at the subject level. Court (1996) reported a statistically significant positive correlation at the institutional level between the 1992 RAE scores and the 1992-95 teaching quality assessment scores in England, and observed that the old universities tended to have better quality ratings for both research and teaching. However, caution is required when making comparisons involving different disciplines. Cluster analysis was used in this study to examine in detail the relationship between quality in research and quality in teaching and learning in just one subject area: business and management studies. Only those departments in England, Scotland and Wales having both teaching quality ratings and 1996 research quality ratings were included in an initial sample of 92 departments (counting as two departments each of the four units of assessment which were given two quality ratings in the 1996 Research Assessment Exercise).

Cluster analysis is helpful in exploring multivariate data, without attempting to prove or refute any preconceived hypotheses. Unlike discriminant analysis, cluster analysis collects similar objects together into new groupings, rather than assigning them to groups which have already been defined. (For a more detailed description of cluster analysis see, for example, Everitt, 1993, Kaufman and Rousseeuw, 1990, and SPSS, 1994.)

An agglomerative hierarchical approach was used to gain an initial understanding of the most meaningful number of clusters of departments of business and management studies with which to proceed. Agglomerative methods start with all objects apart. At each step, the two most similar clusters are merged, until only one remains. After some experimentation, the following variables were chosen as the basis for further investigation: 1996 research quality ratings (using a seven-point scale, converting 3b to 3, 3a to 4, 4 to 5, 5 to 6, and 5* to 7); 1996 research-active faculty FTEs; 1996 percentages of faculty active in research; and teaching quality assessment ratings (converting "satisfactory" to 1, "highly satisfactory" to 2, and "excellent" to 3). Quality ratings are ordinal variables; faculty numbers are numeric variables; and, by convention, the number of faculty active in research is expressed as a percentage of the total. These different types of variables were converted to a common scale by standardizing each variable to a range of between 0 and 1 in order to combine them in the same analysis.

An icicle plot was produced, illustrating the last 20 stages at which of the 92 departments were joined together (Figure 8). At stage 73, there are 20 clusters of similar departments, each represented by a shaded "icicle" at the bottom of the plot. At stage 74, the two most similar clusters are joined together, reducing the number of separate clusters by one. This process continues (working upwards) until at the 92nd step, all 92 departments form one cluster.

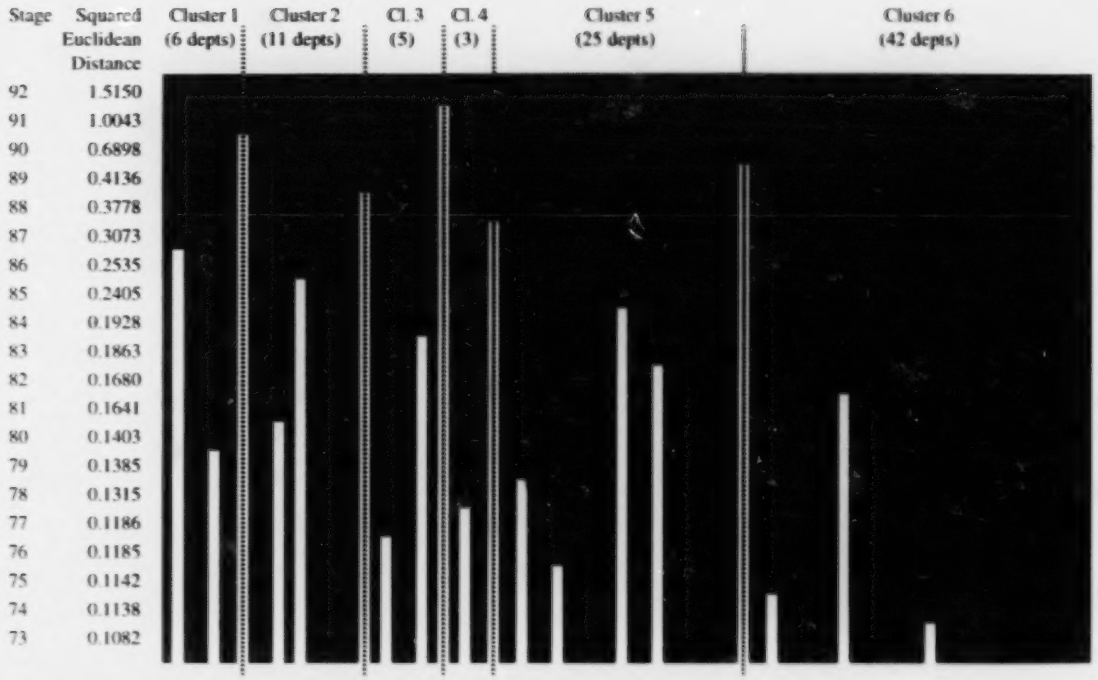

Fig. 8. Icicle Plot of the Last 20 Stages in Hierarchical Agglomeration

The squared Euclidean distance between the clusters being merged at each stage of the calculation was computed. This distance increases as progressively more dissimilar clusters are combined. In this particular example, the squared Euclidean distance between the clusters starts to increase relatively steeply from stage 87 onwards, suggesting that working with perhaps five or six clusters would give rise to the most useful interpretation of the results in practice.

Agglomerative clustering techniques such as this suffer from the potential disadvantage that it is impossible to undo what has already been done in a previous step in the calculation. Once two objects have been joined together, they cannot subsequently be separated, and the end result may be sub-optimal. In general, it may be preferable to use an iterative technique, such as a *k*-means cluster analysis, designed to find a partition having objects in the same cluster being close or related to each other, and with objects in different clusters being far apart or very different. After establishing that it would be appropriate to form alternatively five or six groupings of departments, *k*-means cluster analyses were performed, using the transformed variables. Using a six-cluster analysis produced the final cluster centers after iteration shown in Table 2.

Table. 2. K-Means Cluster Analysis, using Six Clusters—Cluster Centers, the Number of Departments in Each Cluster, and the Number of these Departments in Old Universities

Cluster	Teaching Quality Rating	Research Quality Rating	Research-Active FTEs	Research-Active Faculty %	Total No. of Depts.	Depts. In Old Univs.
1	1.0000	0.9583	0.7345	0.9714	4	4 (100%)
2	1.0000	0.6667	0.3605	0.7943	10	8 (80%)
3	0.0882	0.6667	0.2913	0.8790	17	17 (100%)
4	1.0000	0.3333	0.1317	0.1959	7	1 (14%)
5	0.0000	0.2963	0.1315	0.6254	18	10 (56%)
6	0.0139	0.1852	0.1087	0.1460	36	1 (3%)

The six-cluster groupings suggest that there is a group (cluster 1) of 4 departments which have both the best research ratings and consistently "excellent" teaching ratings. They are the "premier league," or the institutions which Astin and Chang (1995) defined as "high-high." They are the largest departments (in terms of research-active FTEs), and have the highest percentages of research-active faculty.

There are two groups of departments just below the premier league in terms of research quality. One group (cluster 2, 10 departments) has uniformly "excellent" teaching: the other (cluster 3, 17 departments) has generally "satisfactory" teaching quality. Apart from this, there is little to distinguish between these two groups: they are medium-sized with high percentages of research-active faculty.

There are a further two groups of departments below these in research quality. Cluster 4 (7 departments) has uniformly "excellent" teaching, coupled with a low percentage of research-active faculty. On the other hand, cluster 5 (18 departments) has uniformly "satisfactory" teaching, with a quite high percentage of research-active faculty. Departments in cluster 4 are likely to be oriented primarily towards teaching, but having small percentages of faculty active in research has met with a certain degree of success. Departments in cluster 5 are research-oriented, judging by their fairly high percentages of research-active faculty, but are apparently excelling in neither research nor teaching.

Finally, there is a group of 36 departments which are generally small, with low percentages of research-active faculty, which have generally the lowest research ratings, and "satisfactory" teaching quality.

The cluster analysis solutions are not unique, and care should be taken to avoid overinterpretation of the results. In the present example, it may be noted that the numbers of institutions in the six-cluster solutions differ between the agglomerative technique (Figure 8) and the iterative, *k*-means cluster analysis (Table 2). A somewhat different grouping was also obtained from a 5-cluster analysis. Nonetheless, a comparison of the results of the cluster analysis with those reported in other studies indicates the validity of several general conclusions.

BEST COPY AVAILABLE

Clearly, there is no consistent connection between high quality in research and high quality in teaching, at least in business and management studies. This supports the view (also cited by Gibbs, 1995a) that it is one of the "myths of higher education" that good researchers are good teachers (Terenzini and Pascarella, 1994). The proportion of departments in old universities in each of the clusters shown in Table 2 is consistent, in the case of business and management studies, with Court's general contention that quality in research and quality in teaching and learning tend to be concentrated in the old universities (Court, 1996). However, there are also clusters demonstrating notable exceptions to this observation, particularly cluster 5, which includes ten departments in old universities which do not have high ratings in either teaching or research. The results of the cluster analysis are also consistent with the earlier conclusion that research quality and size (as indicated by the number of research-active faculty FTEs) are positively correlated.

Cluster analysis can be a useful alternative to ranking methods. Ranks tend to be particularly sensitive to sampling variability, and there is in general no straightforward way to place interval estimates around institutional rankings (Goldstein and Spiegelhalter, 1996). Knowing, however, that a department falls within a cluster of departments which share similar attributes, and that there are other clusters of departments which share different characteristics, may be more helpful in policy formulation than knowing a precise numerical ranking which has little intrinsic meaning. It will be of interest in future studies to apply the cluster analysis technique with other data from sources such as the NRC research-doctorate study or publishers' rankings.

It is of interest to compare these results with a U.S. study conducted to determine whether some colleges and universities succeed in emphasizing both teaching and research (Astin and Chang, 1995). This study found that no institutions (in a sample of 212) had both a very strong Research Orientation and a very strong Student Orientation (with both types of orientation being determined by faculty survey responses). With less stringent criteria, eleven institutions, all private residential liberal arts colleges, were identified as high in both Research and Student Orientations. By comparison, all of the members of a group with strong Research Orientation and weak Student Orientation were research universities. Thus clusters of institutions with similar characteristics emerged. While the identifying characteristics in the U.S. study may be different from those of the U.K. analysis, the results are consistent in identifying clusters and in suggesting that only a small number of business and management studies departments (U.K.) or institutions (U.S.) are strong in both research and teaching.

Implications for the Shaping of Higher Education

In higher education, as in other fields of human endeavor, the use of particular performance indicators inevitably affects the behavior of those whose efforts are being measured—for better or for worse, either intentionally or otherwise. The measurement of the quality of teaching and learning in the U.K. has undoubtedly raised awareness of these issues, and to this extent, it has unquestionably been beneficial. The measures are important for external accountability, and hopefully lead to program improvements. It has been noted, however, that "... quality assessment is a necessary but insufficient condition for quality enhancement" (Yorke, 1996). There continue to be

perceptions of declining standards of scholastic achievement. The Department for Education and Employment has said, "It is not clear whether the objectives set by universities are sufficiently demanding and whether the standards attained by graduates meet the requirements of the modern world" (Department for Education and Employment [DfEE], 1997). Teaching quality assessment as practiced so far has been mission-dependent, and has been concerned more with procedures and processes, while making no attempt to benchmark the quality of graduates produced against absolute standards (Midwinter, 1997). The National Academies Policy Advisory Group has observed that, "Tension exists between the public rhetoric about the preservation of quality in universities and the reality of the continuing deterioration in the units of resource for both teaching and research" (National Academies Policy Advisory Group, 1996).

The extent to which such challenges can be addressed by placing perhaps greater financial control in the hands of the customers of the system—the students themselves, and their prospective employers—while perhaps also placing greater reliance on an "inspectorate," such as the new Quality Assurance Agency, remains to be seen. Customer choice, better informed though the publication of the outcomes of Teaching Quality Assurance exercises, for example, could radically change the shape of the sector in the future.

The Research Assessment Exercises (RAEs) clearly influence the shaping of the higher education sector to a very considerable extent, primarily because of the associated selectivity in research funding. The criteria for success defined in the RAEs are bound to significantly affect how institutions plan and carry out their research. RAEs are perceived as placing heavy emphasis on academic publications in peer-reviewed journals as a performance measure. The increases in the quality ratings from 1992 to 1996, for example, suggest that institutions may be increasing their efforts in this area. However, there is a danger that departments will adopt increasingly narrow and inappropriate definitions of research, losing touch with the tripartite role identified for them in the Technology Foresight agenda (Parliamentary Office of Science and Technology, 1993): to educate and train qualified manpower; to undertake research and add to knowledge; and to make the knowledge base accessible. There is also a danger that departments will be constrained by the time horizons imposed by periodic RAEs, militating against more speculative, long-term research which, if not done by the universities, is not likely to be done at all.

Additional implications for the shaping of higher education may result from correlations between research quality (as measured by Research Assessment Exercises) and teaching quality (as measured in the Teaching Quality Assessments), at the institutional level (Court, 1996). These may simply be the result of different overall levels of funding in the different parts of the sector, or there may some direct causal relationship between the two. There has been much debate over this. (See, for example, Barnett, 1990 and Gibbs, 1995a.) It is important to use performance indicators which capture all of the issues involved, because policies based only on a narrow range of performance indicators can have unintended side-effects. For example, redoubled efforts in the U.K. to concentrate funding on a few elite universities, highly esteemed for their RAE quality ratings, could have adverse and unintended effects on teaching quality elsewhere in the higher education sector. Gibbs (1995b) has suggested that increased efforts to obtain higher research ratings have already led to the abandonment of innovations in teaching.

The U.K. Research Assessment Exercises and the U.S. study of research-doctorate programs clearly differed in goals, methodology, and use. The broad purpose in the U.K. was to focus public funding on centers of research excellence, in order to use scarce resources to best effect. The goals of the 1993 U.S. study were to update a similar 1982 assessment, to explore the feasibility of replicating or improving objective measures of program quality, to compare data from the 1982 and 1993 studies, to create a data base for further analysis, and to make the findings accessible to educators, administrators, students and policymakers. It was hoped that the study would assist students in the selection of research-doctorate programs, inform the practical judgment of decisionmakers, and provide a large database for use by scholars in the study of higher education. Despite the differences in goals, the two assessments have resulted in similar products—rankings of the quality of research activities within departments in higher education institutions. With appropriate analysis of the results, each provides a clustering of departments with similar characteristics and comparable measures of quality.

Teaching quality assessments in the U.K. and the U.S. differ more significantly, since the U.S. does not have a generally applied method of rating teaching quality. However, the fundamental issues are similar. Policymakers and educators in both countries are concerned with student academic achievement, with teaching effectiveness, and with the resources available for both teaching and research. Questions of the relationship between research quality and teaching quality, and of the appropriate emphasis to be placed on research and teaching in promotion decisions, are debated on both sides of the Atlantic. These and other debates could be better informed by greater international awareness of similar activities in different frameworks. Once past the sometimes daunting superficial differences in educational systems and processes, an improved understanding of similarities can promote quality enhancements and greater understanding within and among nations.

References

Astin, A. W. and Chang, M. J. (1995). Colleges that emphasize research and teaching. Change, 27 (5), 45-49.

Barnett, R. (1990). The Idea of Higher Education. Buckingham: The Society for Research into Higher Education and Open University Press.

Bensman, S., Hamaker, C. A. and Wilder, S. J. (1997). The Louisiana State University serials redesign project: an exercise in optimization. Manuscript in preparation.

Court, S., (1996, April). The Research Assessment Exercise: The Basis for An Unofficial Gold Standard in Higher Education? Paper presented to "The Dilemmas of Mass Higher Education" conference at Staffordshire University at Stoke-on-Trent.

Department for Education and Employment (1997). The Size and Structure of Higher Education. Evidence to the National Committee of Inquiry into Higher Education (the "Dearing Committee"), January 1997. London: Author.

Everitt, B. S. (1993). Cluster Analysis. New York, Toronto: John Wiley & Sons Inc.

Fogel, D. (1996). Uses of the NRC Assessment of Research-Doctorate Programs. Presentation in panel discussion at the 36th Annual Forum of the Association for Institutional Research, Albuquerque, New Mexico.

Gibbs, G. (1995a). The relationship between quality in research and quality in teaching. Quality in Higher Education, 1(2), 147-157.

Gibbs, G. (1995b). Promoting excellent teaching is harder than you'd think. Change, 27 (3), 17-20.

Goldberger, M. L., Maher, B. A., and Flattau, P. E. (Eds.). (1995). Research-Doctorate Programs in the United States: Continuity and Change. Washington, D.C.: National Academy Press.

Goldstein, H. and Spiegelhalter, D. J. (1996). League tables and their limitations: statistical issues in comparisons of institutional performance. Journal of the Royal Statistical Society, Series A, 159, Part 3, 385-443.

Graham, H. D. and Diamond, N. (1997). The Rise of American Research Universities: Elites and Challengers in the Postwar Era. Baltimore and London: The Johns Hopkins University Press.

Higher Education Funding Council for England (1994). The Quality Assessment Method from April 1995. Circular 39/94. Bristol: Author.

Higher Education Funding Council for England (1995). Report on Quality Assessment 1992-95. Bristol: Author.

Higher Education Funding Council for England (1997). Report on Quality Assessment 1995-96. Bristol: Author.

Higher Education Funding Councils (1994a). 1996 Research Assessment Exercise. Circular RAE96 1/94. Bristol: Authors.

Higher Education Funding Councils (1994b). 1996 Research Assessment Exercise: Units of Assessment. Circular RAE96 3/94. Bristol: Authors.

Higher Education Funding Councils (1995). 1996 Research Assessment Exercise: Criteria for Assessment. Circular RAE96 3/95. Bristol: Authors.

Higher Education Funding Councils (1996). 1996 Research Assessment Exercise: The Outcome. Circular RAE96 1/96. Bristol: Authors.

Joint Planning Group for Quality Assurance in Higher Education (1996a). First Report. April 1996. Bristol: Higher Education Funding Council for England.

Joint Planning Group for Quality Assurance in Higher Education (1996b). Final Report. September 1996. Bristol: Higher Education Funding Council for England.

Jones, L. V., Lindzey, G. and Coggeshall, P. E. (1982). An Assessment of Research-Doctorate Programs in the United States. Washington, D.C.: National Academy Press.

Kaufman, L. and Rousseeuw, P. J. (1990). Finding Groups in Data: An Introduction to Cluster Analysis. New York, Chichester, Brisbane, Toronto, Singapore: John Wiley & Sons Inc.

Maher, B. A. (1996). The NRC's report on research-doctorate programs: its uses and misuses. Change, 28 (6), 54-59.

Midwinter, J. E. (1997). A question of quality. IEE (Institution of Electrical Engineers) Review, March 1997, 66-70.

National Academies Policy Advisory Group (1996). Research Capability of the University System. London: Author.

North Central Association (1994). Handbook of Accreditation 1994-96. Chicago: Author.

Parliamentary Office of Science and Technology (1993). Realising Our Potential: A Strategy for Science, Engineering and Technology. London: Her Majesty's Stationery Office.

Patrick, W. J. and Stanley, E. C. (1996). Assessment of research quality. Research in Higher Education, 37(1), 23-42.

Scottish Higher Education Funding Council (1996). Quality Assessment 1996-97: The SHEFC Approach. Edinburgh: Author.

SPSS Inc. (1994). SPSS Professional Statistics™ 6.1. Chicago: Author.

Taylor, J. (1995). A statistical analysis of the 1992 research assessment exercise. Journal of the Royal Statistical Society, Series A 158 (Part 2): 241-261.

Terenzini, P. T. and Pascarella, E. T. (1994). Living with myths: undergraduate education in America. Change, 26 (1), 28-32.

Universities Funding Council (1992). Research Assessment Exercise 1992: The Outcome. Circular letter 26/92. Bristol: Author.

Universities Statistical Record (1995). University Statistics 1993-1994: Volume Three: Finance. Cheltenham: Universities and Colleges Admissions Service (USR Publications).

Webster, D. S. and Skinner, T. (1996a). Rating PhD programs: what the NRC report says..and doesn't say. Change, 28 (3), 22-44.

Webster, D. S. and Skinner, T. (1996b). Webster and Skinner respond. Change, 28 (6), 59.

Yorke, M. (1996). Shouldn't Quality be Enhanced, Rather than Assessed? Tertiary Education And Management, 2(1), 86-94.